On Central Limit Theorems In Geometrical Probability

Florin Avram
Dimitris Bertsimas

WP# 3312-MS

July 14, 1991

On Central Limit Theorems
In Geometrical Probability

Florin Avram
Dimitris Bertsimas

WP# 3312-MS July 14, 1991

On central limit theorems in geometrical probability

Florin Avram * Dimitris Bertsimas †

July 14, 1991

Abstract

We prove central limit theorems (CLT) for the following problems in geometrical probability when points are generated in the $[0,1]^d$ cube according to a Poisson point process with parameter n:

1. The length of the kth-nearest graph $N_{k,n}$, in which each point is connected to its k-th nearest neighbor.
2. The length of the Delaunay triangulation D_n of the points.
3. The length of the Voronoi diagram V_n of the points.

We show using the technique of dependency graphs of Baldi and Rinot that the dependence range in all these problems converges quickly to 0 with high probability. In addition, our approach leads to more efficient sequential and parallel algorithms for this class of problems on randomly distributed points.

1 Introduction

In a pioneering paper by Beardwood, Halton and Hammersley [4] and continued in Steele [11] it was shown that the lengths L_n of several combinatorial optimization problems (the

*Florin Avram, Department of Mathematics, Northeastern University. The research of the author was partially supported from a grant from Northeastern University.

†Dimitris Bertsimas, Sloan School of Management, MIT. The research of the author was partially supported by the National Science Foundation under grants DDM-9014751, DDM-9010332 and by a Presidential Young Investigator Award with matching funds from Draper Laboratories.

traveling salesman, minimum matching, minimum spanning tree, minimum Steiner tree, etc.) satisfy laws of large numbers, when their input consists of n random iid points X_k, $k = 1, \ldots, n$ in the cube $[0,1]^d$.

It is also believed (see for example Steele [12]), but yet unknown, that they satisfy central limit theorems (CLT), since, while the edges of the above optimal graphs are not independent, the dependence "seems to be local". In trying to make this intuitive idea precise we succeeded in proving CLTs for three graphs, which are among the most fundamental constructions in computational geometry (see for example, Preparata and Shamos [9]):

1. The length $N_{k,n}$ of the kth-nearest graph, in which each point is connected to its k-th nearest neighbor.

2. The length V_n of the Voronoi diagram of the points, which is the collection of Voronoi polygons for each point. Given a collection of points X, a Voronoi polygon around a point O is the set of all points in the plane which are closer to O than to any other point in X (see figure 1).

3. The length D_n of the Delaunay triangulation of the points, which is the graph defined on the given points with an edge between them if they are Voronoi neighbors (see figure 1). This graph is indeed a triangulation, i.e., a subdivision of the convex hull of points in triangles.

Both the Voronoi diagram and the Delaunay triangulation are fundamental constructions in computational geometry, since many algorithms for solving geometrical problems are based on their construction. In particular, the Delaunay triangulation is quite important for the MST, since it contains the MST as a subgraph, i.e., points which are neighbors in the MST have to be also Voronoi neighbors (see figure 1). This property leads to an efficient algorithm for the MST, since one first constructs the Delaunay triangulation in $O(n \log n)$ time and then runs the greedy algorithm on its graph, leading to an $O(n \log n)$ algorithm for the MST as well.

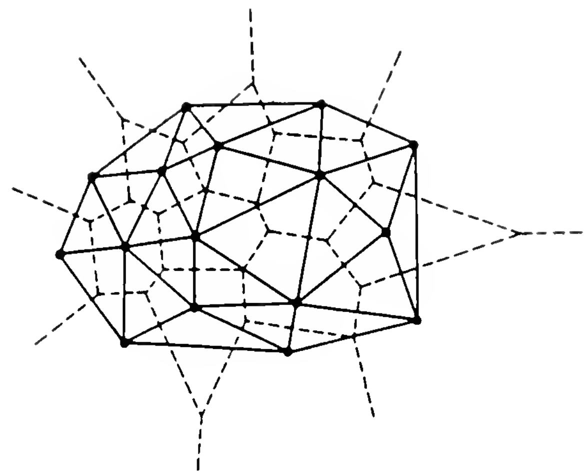

Figure 1: The Voronoi diagram and the Delaunay triangulation

We believe that our results give some partial insight on why a CLT might hold for the MST as well. Ramey [10] has attempted to prove a CLT for the MST, but his approach, although very interesting, did not succeed since he needed some unproven, but plausible, lemmas from continuous percolation.

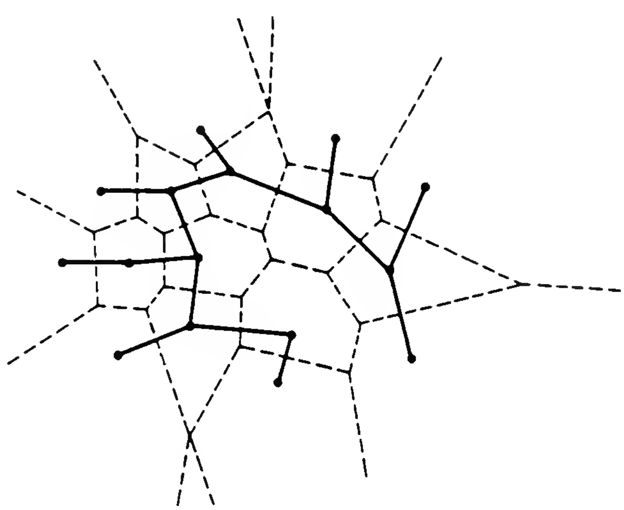

Figure 2: The MST is a subgraph of the Delaunay triangulation

The CLT for $N_{1,n}$ has already been obtained by Bickel and Breiman [5], who used very complicated fourth moment estimates. They wrote: "Our proof is long. We believe that this

is due to the complexity of the problem. Nearest neighbor distances are not independent". We used instead a simple geometrical approach, which works very well in all these problems and it has the potential to work for other problems also.

Our results are established for the Poisson model, which is asymptotically equivalent to the usual Euclidean model, but more convenient to work with.

Poisson model

Let X_i, $i = 1, \ldots, N_n$ be the points of a Poisson process with intensity n on R^d, which lie in a d-dimensional cube, so that N_n is a Poisson random variable with mean n.

The paper is structured as follows. Section 2 contains the proof of our main result that for all three problems:

$$\lim_{n\to\infty} Pr\{\frac{L_n - E[L_n]}{\sqrt{Var[L_n]}} \le x\} = \Phi(x), \tag{1}$$

where $\Phi(x)$ is the cumulative distribution of a standard $N(0,1)$ normal. In Section 3, we show that the lengths $E[L_n]$ of the k-th nearest neighbor graph,the Delaunay triangulation and the Voronoi diagram satisfy:

$$\lim_{n\to\infty} \frac{E[L_n]}{n^{\frac{d-1}{d}}} = \beta_d. \tag{2}$$

The constants β_d appearing in (2) have been explicitly computed by Miles [7], which is in sharp contrast with the problems studied in [4], [11] with the exception of the MST constant that we calculated recently in [1]. We include for convenience simple direct computations for the constants. Finally in the final section we remark that our approach leads to more efficient sequential and parallel algorithms for this class of problems on randomly distributed points.

2 Central limit theorems

In this section we prove CLTs for the 3 problems we consider. We first outline the methodology and comment on its applicability to other combinatorial problems.

2.1 Intuitive idea

We identify a certain geometrical configuration A_n, such that $Pr\{A_n\} \rightarrow 1$, whose occurrence implies independence of the configuration at points which are far enough. More precisely, there is a cutoff distance, such that if the event A_n happened, then deterministically, the configuration around a given point is not influenced by that of points further than the cutoff distance.

For all the three problems, the event A_n is the event that, in the subdivision of $[0,1]^d$ into $p_n = \frac{n}{\log n}$ equal subcubes, each subcube contains at least one and at most $e \log n$ of the points X_k. If all the neighboring subcubes around a point are nonempty, we then show that only a finite number of them determines the k-th nearest neighbors or the Voronoi neighbors of a point. Thus, conditional on the event A_n, the three problems exhibit "m-dependence" (at the subdivision level) for some finite m. Applying the theory of dependency graphs from [2] yields the CLTs.

As mentioned in the introduction, our approach was inspired by the Ph.D thesis of Ramey [10], who attempted a similar approach for the minimum spanning tree. In order to see how the approach we described for the three problems we considered might generalize for the MST, consider the event $B_{n,k}$ that there exist 2 points which are Voronoi neighbors and lie at some distance l of each other, but can also be connected by a chain of edges which are all shorter than l and the shortest such chain between the two points uses more than k edges for some large k. Such two points will not be neighbors in the MST and the decision not to connect them *is affected* by the position of far away points. Let $B'_{n,k}$ be the complement of the event $B_{n,k}$

Conditioning in this case on $A_n \cap B'_{n,k}$ makes the problem m-dependent with m finite and thus showing that $Pr\{B_{n,k}\} \rightarrow 0$, as $n \rightarrow \infty$ would establish the CLT for the MST (modulo some technicalities). Further analogies with continuous percolation, when $d = 2$, make the above quite plausible, even though a rigorous proof has not been submitted.

To summarize, in the nearest neighbor, the Voronoi diagram and the Delaunay triangulation, we have established the geometrical configuration whose probability of occurrence

tends to 1 and which implies finite range dependence and thus the CLT. In contrast, for the MST we know the geometrical configuration, but we can not show its probability of occurrence tends to 1. Finally, in the TSP and minimum matching problems we do not even have a clear geometric argument of why dependence is local.

2.2 Dependency graphs

The fundamental concept that captures the idea of local dependence is that of dependency graphs. Introduced in Petrovskaya and Leontovitch [8] and applied to several problems by Baldi and Rinott [2], [3] dependency graphs are defined as follows:

Let X_a, $a \in V$ be a collection of random variables. The graph $G = (V, E)$ is said to be a *dependency graph* for X_a, if for any pair of disjoint sets A_1, $A_2 \in V$ such that no edge in E has one endpoint in A_1 and the other in A_2, the sets of random variables X_a, $a \in A_1$ and X_a, $a \in A_2$ are independent. The following is the fundamental theorem that captures local dependence.

Theorem 1 *([3]) Let $\{X_{a,n},\ a \in V_n\}$ be random variables having a dependency graph $G_n = (V_n, E_n)$, $n \geq 1$. Let $S_n = \sum_{a \in V_n} X_{an}, \sigma_n^2 = Var[S_n] < \infty$. Let D_n denote the maximum degree of G_n and suppose $|X_{an}| \leq B_n$ almost surely. Then*

$$|Pr\{\frac{S_n - E[S_n]}{\sigma_n} \leq x\} - \Phi(x)| \leq 32(1+\sqrt{6})(\frac{|V_n| D_n^2 B_n^3}{\sigma_n^3})^{\frac{1}{2}}. \tag{3}$$

Thus if $\frac{|V_n| D_n^2 B_n^3}{\sigma_n^3} \to 0$ as $n \to \infty$

$$\frac{S_n - E[S_n]}{\sigma_n} \to \mathcal{N}(0,1). \tag{4}$$

Remark:

The above theorem quantifies the fact that enough independence (D_n small), enough boundness (B_n small) and enough variability (σ_n large) imply the CLT.

2.3 The CLT

Let L_n be the length of the graphs $N_{k,n}, D_n$ and V_n. We subdivide the cube $[0,1]^d$ into a set C of m^d subcubes with $m^d = (\frac{n}{\log n})$. The subcubes will be denoted by indices $\vec{i} = (i_1, \ldots, i_d)$, where each i_j takes values from 1 to m.

We decompose the total length L_n as

$$L_n = \sum_{\vec{i} \in C} L_{\vec{i},n} \tag{5}$$

where $L_{\vec{i},n}$ is the sum of all edges with both ends in the subcube $\vec{i}$, plus the sum of the portion of the edges within the subcube with only one endpoint in the subcube $\vec{i}$. Let $N_{\vec{i},n}$ be the number of points falling in the cube $\vec{i}$ and A_n be the event that the subcubes will be non-empty and contain at most $e \log n$ points, i.e.,

$$A_n = \cap_{\vec{i} \in C}\{1 \leq N_{\vec{i},n} \leq e \log n\}. \tag{6}$$

We first show that

Lemma 2

$$\lim_{n \to \infty} Pr\{A_n\} = 1.$$

Proof

The random variable $N_{\vec{i},n}$ is Poisson with parameter $\lambda = \frac{n}{m^d} = \log n$. Thus

$$Pr\{1 \leq N_{\vec{i},n} < e\lambda\} = 1 - Pr\{N_{\vec{i},n} = 0\} - Pr\{N_{\vec{i},n} \geq e\lambda\}.$$

But $Pr\{N_{\vec{i},n} = 0\} = e^{-\lambda}$ and

$$Pr\{N_{\vec{i},n} \geq e\lambda\} = \sum_{r=e\lambda}^{\infty} e^{-\lambda}\frac{\lambda^r}{r!} \leq e^{-\lambda}\frac{\lambda^{e\lambda}}{(e\lambda)!}[\sum_{k=0}^{\infty}\frac{\lambda^k}{(e\lambda)^k}] =$$

$$\frac{e}{e-1}e^{-\lambda}\frac{\lambda^{e\lambda}}{(e\lambda)!} \leq \frac{\sqrt{e}}{(e-1)\sqrt{2\pi}}\frac{e^{-\lambda}}{\sqrt{\lambda}} \leq e^{-\lambda},$$

where we used Stirling's bounds for the next to last inequality. Therefore,

$$Pr\{1 \leq N_{\vec{i},n} < e\lambda\} \geq 1 - 2e^{-\lambda} = 1 - \frac{2}{n}.$$

Since we are using the Poisson model, the random variables $N_{\vec{i},n}$ are independent and therefore

$$Pr\{A_n\} \geq (1 - \frac{2}{n})^{m^d} = (1 - \frac{2}{n})^{\frac{n}{\log n}}.$$

Taking limits we obtain

$$\lim_{n \to \infty} Pr\{A_n\} \geq \lim_{n \to \infty} e^{-\frac{2}{\log n}} = 1$$

and thus lemma 2 follows. □

We introduce now a distance between subcubes $\vec{i}$ and $\vec{j}$:

$$d(\vec{i}, \vec{j}) = max_{1 \leq r \leq d}\{|i_r - j_r|\}.$$

Let

$$S_{\vec{i},R} = \{\vec{j} \in C : \ d(\vec{i}, \vec{j}) \leq R\},$$

i.e., $S_{\vec{i},R}$ denotes the sphere of subcubes of radius R around $\vec{i}$. Moreover, if A, B are two sets of subcubes, we let

$$d(A, B) = min_{\vec{i} \in A, \vec{j} \in B} d(\vec{i}, \vec{j}).$$

The following proposition is the heart of our development of the CLT and captures the idea of "local dependence" in the three problems we consider.

Proposition 3 *Conditional on the event A_n having occurred, there exists a fixed number R (depending only on the dimension d) such that if A, B is any pair of sets of subcubes with $d(A, B) > R$, then the corresponding collections of random variables $\{L_{\vec{i},n}, \ \vec{i} \in A\}$ and $\{L_{\vec{i},n}, \ \vec{i} \in B\}$ are independent.*

Proof

Let $X_{\vec{i}}$ denote the set of random variables $\{X_v, \ v \in \vec{i}\}$, i.e the collection of the locations of the points in a subcube. We note first that the point process on $[0, 1]^d$ obtained from the Poisson point process by conditioning on A_n, will retain the property that $X_{\vec{i}}$ will be independent.

We will prove the proposition for $d = 2$, where the geometry is easier to visualize, but the proof holds with no changes in any d. Similarly, for ease of exposition in the k nearest neighbor graph we present the case $k = 1$.

For the nearest neighbor graph (with $d = 2$ and $k = 1$), we prove the proposition with $R = 4$. Note first that if two points $X_k \in \vec{i}$ and $X_l \in \vec{j}$ are nearest neighbors, then $d(\vec{i}, \vec{j}) \leq 2$, since if $d(\vec{i}, \vec{j}) \geq 3$, then a point in a subcube at distance one of $\vec{i}$ is guaranteed to be closer than X_l (see also figure 3).

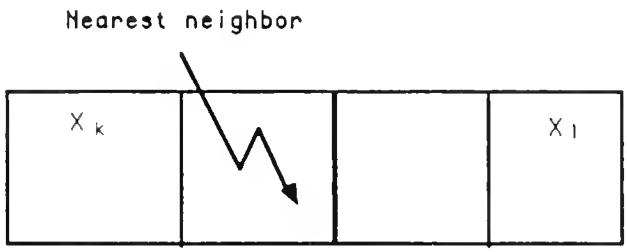

Figure 3: Local dependence in the nearest neighbor graph

Moreover, the decision which of the points X_r belonging in subcubes in the sphere $S_{\vec{i},2}$ is the nearest neighbor of $X_k \in \vec{i}$, is in no way affected by any point X_r outside $S_{\vec{i},2}$. (Note that the analog of this property fails to hold for the MST, matching and TSP). But, given the event A_n, $\{L_{\vec{i},n}, \ \vec{i} \in A\}$ is some function $f\{X_{\vec{i}}, \ \vec{i} \in S_{A,2}\}$ and similarly $\{L_{\vec{i},n}, \ \vec{i} \in B\}$ is some function $f\{X_{\vec{i}}, \ \vec{i} \in S_{B,2}\}$. Since $d(A, B) > 4$ implies that $S_{A,2}$, $S_{B,2}$ are disjoint and since $X_{\vec{i}}$ are independent, the proposition follows.

We now turn our attention to the Voronoi diagram. Arguing as before, the Voronoi polygon of a point $X_k \in \vec{i}$ is included in the sphere $S_{\vec{i},2}$, since the Voronoi polygon of X_k is the set of points which are nearest to the point X_k than any other point X_l. Thus, if $X_k \in \vec{i}$ and $X_l \in \vec{j}$ are Voronoi neighbors then $d(\vec{i}, \vec{j}) \leq 4$, since the midpoint of X_k and X_l belongs to the Voronoi polygon of both. In fact, with more effort we can check that $d(\vec{i}, \vec{j}) \leq 3$, since if $d(\vec{i}, \vec{j}) \geq 4$, then any circle through X_k, X_l has to contain a full subcube $\vec{r}$, where a point X_t lies (since all subcubes are non-empty) and thus X_k, X_l cannot be Voronoi neighbors.

The decision which of the points X_r belonging in subcubes in the sphere $S_{\vec{i},4}$ are Voronoi neighbors is the nearest neighbor of $X_k \in \vec{i}$, is in no way affected by any point X_r outside $S_{\vec{i},4}$. Thus, arguing as in the nearest neighbor case, $R = 8$ is sufficient to prove the proposition. □

Since we want to apply theorem 1 we want an upper bound on the length of $L_{\vec{i},n}$, which is provided by the following proposition.

Proposition 4 *Given the event A_n*

$$L_{\vec{i},n} \leq \frac{e\sqrt{d}(\log n)^{1+\frac{1}{d}}}{n^{\frac{1}{d}}}. \tag{7}$$

Proof

Since every distance is at most the length $\frac{\sqrt{d}}{m}$ of the diagonal of the subcube, we have that

$$L_{\vec{i},n} \leq N_{\vec{i},n}\frac{\sqrt{d}}{m}.$$

Given the event A_n, $N_{\vec{i},n} < e \log n$ and since $m = (\frac{n}{\log n})^{\frac{1}{d}}$ we obtain (7). □

Finally, in order to apply theorem 1 we need a lower bound on the variance of L_n. This is provided in the following proposition.

Proposition 5 *There exists a constant f_d depending only on the dimension such that*

$$Var[L_n] \geq f_d n^{\frac{d-2}{d}}. \tag{8}$$

Proof

We will use a general technique from Ramey [10], which we illustrate for $d = 2$. We subdivide the cube $[0,1]^2$ in n squares of dimension $\frac{1}{\sqrt{n}}$ with area $\frac{1}{n}$ each. We identify a particular configuration, which has a strictly positive probability of occurring and has nonzero variability. We describe the construction for the nearest neighbor graph. Let $\epsilon = \frac{1}{8\sqrt{n}}$. Consider all the subcubes that have the following properties:

1. They contain exactly 2 points inside a circle with center the center of the subcube and radius ϵ, and no other points in the subcube.

2. A ring of points exist near the outside of the boundary at distance at most 2ϵ of another. In particular, we demand that all the 36 smaller subcubes of dimension ϵ that surround the initial subcube are non-empty (see figure 4).

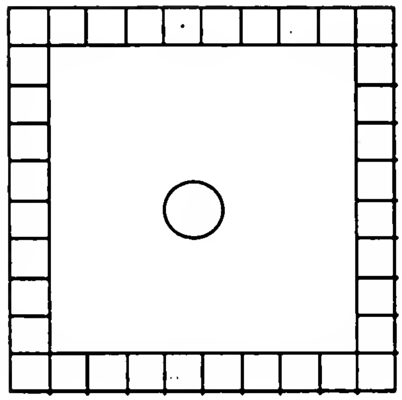

Figure 4: The configuration for the nearest neighbor graph

Note that no point outside such a subcube has a nearest neighbor inside (since its distance to a point inside is at least 3ϵ and there are points outside within 2ϵ). Moreover, the two points inside the circle are nearest neighbors of each other. Their distance d has a certain variance $\frac{\sigma}{n}$ (easily computable). Finally, the probability of a such configuration is $p = e^{-\lambda\pi\epsilon^2}\frac{(\lambda\pi\epsilon^2)^2}{2}e^{-\lambda(\frac{1}{n}-\pi\epsilon^2)}(1-e^{-\lambda\epsilon^2})^{36} = 1.8310^{-76} > 0$.

Let N be the number of subcubes $I = \{i_1, \ldots, i_N\}$ satisfying the two properties above. Then $E[N] = np$. Let F_n be the σ-algebra determined by the random variables $i_1, \ldots, i_N$ and by the positions of all the points lying outside the N subcubes. Then,

$$Var[L_n] = Var[E[L_n|F_n]] + E[E[L_n^2|F_n] - (E[L_n|F_n])^2] =$$

$$Var[E[L_n|F_n]] + E[Var[L_n|F_n]] \geq E[Var[\sum_{i\in I} d_i + \sum_{i\notin I} d_i|F_n]] =$$

$$E[\sum_{i\in I} Var[d_i]] = \frac{\sigma}{n}E[N] = \sigma p > 0,$$

where d_i is the sum of all the nearest neighbor distances of the points inside a subcube i. Letting $f_2 = \sigma p$ proves the proposition for the nearest neighbor graph.

For the k-nearest graph the construction is similar except we require that there are $k+1$ points inside the circle, so that all k-nearest neighbors are inside the circle and its of the outside subcubes contains $k+1$ points. .

For the Voronoi diagram and the Delaunay triangulation we consider all subcubes with the following properties: They contain exactly one point inside a circle with center the center of the subcube and radius $\frac{\epsilon}{2}$ and exactly 3 points inside a circle ϵ and no other points in the subcube. The three points in the second circle are located as follows: We subdivide the circle in six equal circular regions. The 3 points are located in alternate regions. As in the nearest neighbor graph a ring of points exist outside the boundary at distance at most 2ϵ of each other. The probability of such configuration is $p_1 > 0$. Note that the point inside has Voronoi neighbors only the three points in the outside circle. Fixing the position of the three points in the outside circle there is a certain variability $\frac{\sigma_V}{n}$, $\frac{\sigma_D}{n}$ for the Voronoi polygon and its Delaunay triangulation of the point inside, all of which are easily computable. Letting N_1 be the number of subcubes I satisfying the two properties above we let U_n be the $\sigma-$algebra determined by the random variables $i_1, \ldots, i_{N_1}$ and by the positions of all the points lying outside the N subcubes and inside the larger circle. Then,

$$Var[L_n] = Var[E[L_n|U_n]] + E[E[L_n^2|U_n] - (E[L_n|U_n])^2] =$$

$$Var[E[L_n|U_n]] + E[Var[L_n|U_n]] \geq E[Var[\sum_{i\in I} d_i + \sum_{i\notin I} d_i|U_n]] =$$

$$\frac{\sigma_V}{n} E[N] = \sigma_V p_1 > 0,$$

where d_i is the sum of the Voronoi polygon (Delaunay triangulation) of the inside point. □

We now have all the ingredients to prove the CLT.

Theorem 6 *The length L_n of the k-nearest neighbor graph, the Voronoi diagram and the Delaunay triangulation satisfies the CLT*

$$\lim_{n\to\infty} Pr\{\frac{L_n - E[L_n]}{\sqrt{Var[L_n]}} \leq x\} = \Phi(x), \tag{9}$$

where $\Phi(x)$ is the cumulative distribution of a standard normal.

Proof

We first establish that the random variable $\frac{L_n - E[L_n]}{\sqrt{Var[L_n]}}$ given that the event A_n has occurred is asymptotically normal. Indeed, we define a dependency graph in the sense of section 2.2 between the $L_{\vec{i},n}$ by putting an edge between $L_{\vec{i},n}$ and $L_{\vec{j},n}$ if $d(\vec{i},\vec{j}) \le R$, where R is defined in proposition 3. Proposition 3 establishes the correction of this construction. Moreover, the maximal degree D_n of this dependency graph is at most R^d, i.e., finite independent of n. Applying theorem 1 we have that $|V_n| = m^d$ and $D_n \le R^d$ from proposition 3. In addition, $B_n \le \frac{e\sqrt{d}(\log n)^{1+\frac{1}{d}}}{n^{\frac{1}{d}}}$ from proposition 4 and $Var[L_n] \ge f_d n^{\frac{d-2}{d}}$ from proposition 5.

Therefore, since

$$\frac{|V_n| D_n^2 B_n^3}{Var[L_n]^{\frac{3}{2}}} \le \frac{m^d (R^d)^2 [\frac{e\sqrt{d}(\log n)^{1+\frac{1}{d}}}{n^{\frac{1}{d}}}]^3}{(f_d n^{\frac{d-2}{d}})^{\frac{3}{2}}} \le u_d \frac{n^{\frac{3}{2}+\frac{3}{d}}}{m^{2d+3}},$$

where u_d is a constant depending on the dimension only. Since $m^d = \frac{n}{\log n}$ we obtain

$$\frac{|V_n| D_n^2 B_n^3}{Var[L_n]^{\frac{3}{2}}} \le u_d \frac{[\log n]^{2+\frac{3}{d}}}{n^{\frac{1}{2}}} \rightarrow 0,$$

as $n \rightarrow \infty$. Applying now theorem 1, we establish that $\frac{L_n - E[L_n]}{\sqrt{Var[L_n]}}$ given that the event A_n has occurred is asymptotically normal N$(0,1)$.

We now need to show that $U_n = \frac{L_n - E[L_n]}{\sqrt{Var[L_n]}}$ is asymptotically normal. But

$$Pr\{U_n \le x\} = Pr\{U_n \le x | A_n\} Pr\{A_n\} + Pr\{U_n \le x | A_n'\} Pr\{A_n'\}$$

Taking limits and using lemma 2 and the asymptotic normality of U_n given the event A_n we establish the theorem. □

Remark:

We can establish rates of convergence to the normal distribution by using (3). We see that

$$|Pr\{\frac{L_n - E[L_n]}{\sqrt{Var[L_n]}} \le x\} - \Phi(x)| = O(\frac{\log n^{1+\frac{3}{2d}}}{n^{\frac{1}{4}}}).$$

3 Computations of the expectations

In this section we will find exactly $E[L_n]$, i.e., we will compute the constants β_d, appearing in (2), using the toroidal rather than the usual Euclidean distances, so that there are no

boundary effects. However, the boundary effects can be shown to be negligible (Jaillet [6]), and so the same results holds.

The k-th nearest neighbor graph

Consider a given point under the toroidal model. Then the length of the nearest neighbor graph $L_n = nR_n$, where R_n is the distance of a given point to its k-th nearest neighbor. Also,

$$E[R_n] = \int_0^{\sqrt{d}} Pr\{R_n \geq r\} dr =$$

$$\int_0^{\sqrt{d}} \sum_{j=0}^{k-1} \frac{e^{-n\, c_d\, r^d}[n\ c_d\ r^d]^j}{j!} dr,$$

where c_d is the volume of the unit sphere in dimension d.

Letting $u = n\ c_d\ r^d$ we obtain that

$$\frac{E[L_n]}{n^{\frac{d-1}{d}}} = \sum_{j=0}^{k-1} \frac{1}{dc_d^{1/d} j!} \int_0^{n\, c_d\, d^{\frac{d}{2}}} e^{-u} u^{j+\frac{1}{d}-1} du.$$

Letting $n \to \infty$ we obtain that

$$\lim_{n\to\infty} \frac{E[L_n]}{n^{\frac{d-1}{d}}} = \frac{1}{dc_d^{1/d}} \sum_{j=1}^{k} \frac{\Gamma(j+\frac{1}{d}-1)}{(j-1)!}. \tag{10}$$

The length of the Voronoi diagram

Let R_n be the length of the perimeter of the Voronoi polygon around a given point O. Then $E[L_n] = n\frac{E[R_n]}{2}$. We will perform the calculation for $d = 2$. Let $E[h_n]$ be the expected contribution to R_n from a point C of the Poisson process. Then $E[R_n] = nE[h_n]$. Conditioning on the distance $OC = r$ (see figure 5), let $E[h_n(r)]$ be the expected contribution to $E[h_n]$ given that $OC = r$. Then,

$$E[h_n] = \int_0^{\sqrt{2}} E[h_n(r)] 2\pi r dr.$$

Moreover, $E[h_n(r)] = 2\int_0^{\sqrt{2}} Pr\{$the length dl at distance l from A is part of the Voronoi diagram$\}dl$. But, $Pr\{$the length dl at distance l from A is part of the Voronoi diagram$\}$ = $Pr\{$no perpendicular bisector cuts the segment $OB\}$ = $Pr\{$no point of the Poisson process belongs in the circle with center B and radius $BO\} = e^{-\pi((\frac{r}{2})^2+l^2)n}$. Thus

$$E[h_n] = 4\pi \int_0^{\sqrt{2}} \int_0^{\sqrt{2}} re^{-\pi((\frac{r}{2})^2+l^2)n} dl dr.$$

Computing the integral we obtain

$$\lim_{n\to\infty} E[h_n]n^{\frac{3}{2}} = 4$$

which leads to

$$\lim_{n\to\infty} \frac{E[L_n]}{\sqrt{n}} = 2.$$

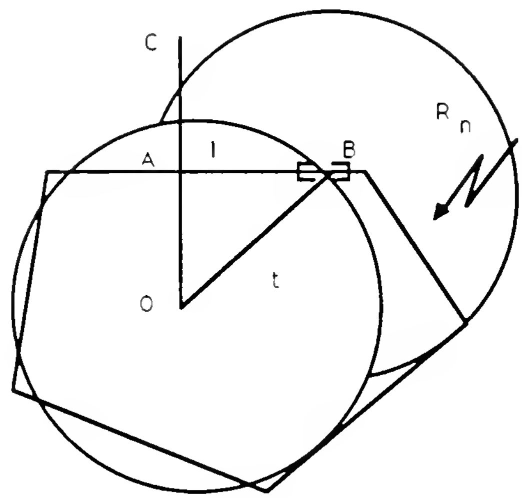

Figure 5: The expectation of the length of the Voronoi diagram

The Delaunay triangulation

Let R_n be the sum of the distances in the Delaunay triangulation from a given point O to its Voronoi neighbors. Then, the expected length of the triangulation is

$$E[L_n] = \frac{nE[R_n]}{2}.$$

Let $h(r)rdr\ d\theta = Pr\{A$ is a Voronoi neighbor of O— A is a point in $rdr\ d\theta\}$. Then

$$E[R_n] = \int_0^{2\pi} \int_0^{\sqrt{2}} rh(r)rdrd\theta.$$

In order to compute $h(r)$ we consider the circle passing through O, A and a third point of the Poisson process B, which is to the right of the segment OA with the property that it contains no other point of the process. It is well known (see for example [9]) that O, A are

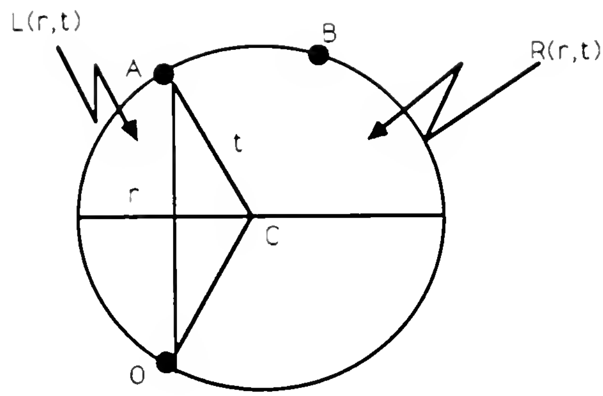

Figure 6: The expectation of the length of the Delaunay triangulation

Voronoi neighbors if and only if no points of the Poisson process lie in the segment of the circle to the left of OA (see figure 6).

Let T_r the radius of the circle and $L(r,t)$, $R(r,t)$ the areas shown in figure 6. Therefore,

$$h(r) = 2\int_0^{\sqrt{2}} e^{-L(r,t)n} dPr\{T_r \leq t\}$$

where we multiplied by 2 because we considered the point C to the right of OA, where we could have considered the case in which C is on the left. Moreover,

$$Pr\{T_r \leq t\} = e^{-R(r,t)n},$$

is the area of the right segment of the circle. If $\phi = sin^{-1}\frac{r}{2t}$ is the angle ACB, then from simple geometry $L(r,t) = \phi t^2 - \frac{rtsin\phi}{2}$ and $R(r,t) = (\pi - \phi)t^2 + \frac{rtsin\phi}{2}$. Computing the integrals we find that

$$\lim_{n\to\infty} E[R_n]\sqrt{n} = \frac{64}{3\pi},$$

which leads to

$$\lim_{n\to\infty} \frac{E[L_n]}{\sqrt{n}} = \frac{32}{3\pi}.$$

4 On the relation of CLTs to efficient algorithms

The approach we used leads also to a faster algorithm to compute the Voronoi diagram and the Delanay triangulation for randomly distributed points. From proposition 3, in order to find the portion of the Voronoi diagram inside a subcube, we need only to consider only those subcubes that are at distance at most $R = 4$ from this subcube. Thus, if the event A_n occurs, there are $O(\log n)$ points that contribute to the portion of the Voronoi diagram inside a subcube. Running an $O(k \log k)$ algorithm with $k = O(\log n)$ gives an $O(\log n \log\log n)$ algorithm to find the Voronoi diagram inside a subcube. Since there are $\frac{n}{\log n}$ subcubes, we obtain an $O(n \log\log n)$ algorithm. If the event A_n does not occur, we simply run any $O(n \log n)$ algorithm. Since the probability of A_n goes to 1, this algorithm runs in time $O(n \log\log n)$ with probability 1. In addition, since this computation can be done in parallel, the above appraoch leads to an $O(\log n \log\log n)$ parallel algorithm with $\frac{n}{\log n}$ processors.

In general, if such an approach were to work for other combinatorial problems, we would obtain not only a CLT, but faster sequential as well as parallel algorithms for their solution.

References

[1] F. Avram and D. Bertsimas, (1990), "The minimum spanning tree constant in geometrical probability and under the independent model; a unified approach", to appear in *Annals Appl. Prob.*

[2] P. Baldi, Y. Rinott, (1989), "Asymptotic normality of some graph related statistics", *Jour. Appl. Prob.*, 26, 171-175.

[3] P. Baldi, Y. Rinott, (1989), "On normal approximations of distributions in terms of dependency graphs", *Annals Prob.*, 17, 1646-1650.

[4] J. Beardwood, J.H. Halton and J.M. Hammersley, (1959), "The shortest path through many points", *Proc. Cambridge Philos. Soc.*, 55, 299-327.

[5] P. Bickel and L. Breiman, (1983), "Sums of functions of nearest neighbor distances, moment bounds, limit theorems and goodness of fit tests", *Ann. Prob.*, 11, 185-214.

[6] P. Jaillet, (1990), personal communication.

[7] R. Miles, (1970), "On the homogeneous planar Poisson point process", *Math. Biosciences*, 6, 85-127.

[8] M. Petrovskaya and A. Leontovitch, (1982), "The central limit theorem for a sequence of random variables with a slowly growing number of dependencies", *Theory Prob. Appl.*, 27, 815-825.

[9] F. Preparata and M. Shamos, (1985), *Computational Geometry*, Springer-Verlag, New York.

[10] M. Ramey, (1983), Ph.D. thesis, Department of Statistics, Yale University.

[11] J.M. Steele, (1981) "Subadditive Euclidean Functionals and Nonlinear Growth in Geometric Probability", *Ann. Prob.*, 9, 365-376.

[12] J.M. Steele, (1988) "Growth Rates of Euclidean Minimal Spanning Trees with Power Weighted Edges", Annals of Prob., 1767-1787.

Date Due